DOCUMENT RESUME

ED 414 160 SE 060 722

AUTHOR Miller-Whitehead, Marie
TITLE An Analysis of Science Scale Scores for Grades 2-8 in Tennessee for 1990-1994.
PUB DATE 1997-09-00
NOTE 15p.; For a related study, see SE 060 860.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Accountability; *Achievement Tests; Educational Change; Elementary Education; Finance Reform; Legislation; Mathematics Education; School Effectiveness; *School Statistics; *Science Education; *Scoring; Standards
IDENTIFIERS Tennessee Education Improvement Act 1991

ABSTRACT

The use of statewide tests of student achievement as one component of accountability are certainly not new. An increasing number of states have mandated statewide testing through legislation aimed at tying financial incentives to a variety of accountability indicators including student achievement. These initiatives have generated several side effects, both positive and negative: (1) there has been a renewed interest in research on factors that influence student achievement; (2) the general public, ever-wary of tax increases, has been given a concrete measurement (however controversial) by which to gauge student success; and (3) teachers, administrators, and other professional educators have become increasingly aware of the public policy implications that quantitative data can have on schools, personnel, and school programs. As the result of a school funding equity lawsuit in the state of Tennessee, new legislation mandating revised school funding formulae and accountability procedures was implemented in 1991 for all K-12 public schools. Part of the accountability procedure includes mandated annual testing of all students in grades 2 through 8 in the areas of science, math, language arts, reading, and social studies. The goal of the Tennessee Education Improvement Act (EIA) of 1991 is to reduce variability among scores in school systems across the state regardless of socioeconomic status (equity), and to ensure that all students are progressing (or value-added) from one year to the next in each of the key subject areas. These goals are reflective of the national trend toward increased accountability in education. This research, focusing solely on the area of science, addresses the following questions: (1) is there evidence of more equity and value-added in student scores?; (2) was variability in scores decreasing?; (3) how do scores compare across years and grade levels?; and (4) what are the implications for curriculum and assessment reforms? The data set for this study consisted of scale science scores in 133 Tennessee public schools, grades 2-8, for the years 1990-1994. The null hypothesis of the investigation was that there is no difference in science scale scores across years or grade levels. (Contains 11 references.) (Author/DKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 160

An analysis of science scale scores for grades 2-8 in Tennessee for 1990-1994

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M Miller-Whitehead

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Marie Miller-Whitehead
Education Consultant
September 8, 1997

SE 060722

BEST COPY AVAILABLE

ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of Dr. Ben Brown, director of Accountability Systems for the Tennessee State Department of Education, and of Dr. Scott Snyder of the Department of Educational Research and Accountability, the University of Alabama at Birmingham, who both provided suggestions, advice, and help in obtaining and organizing the data used in this analysis. The statistical analysis and interpretation of results, however, do not necessarily reflect those of anyone other than the author.

Marie Miller-Whitehead
September 8, 1997

Introduction

The use of statewide tests of student achievement as at least one component of accountability are certainly not new. The current national examination of school effectiveness indicators began in 1981 when Secretary of Education Bell created the National Commission on Excellence in Education. The Commission's controversial report, A Nation at Risk, was published in 1983 and resulted in widespread interest in national standards for education and in increased accountability for educators at all levels (Gardner et al.). Most school districts provide for periodic standardized testing of students if not on an annual basis at least at certain grade levels, such as fourth, eighth, and eleventh. However, an increasing number of states have mandated statewide testing through legislation aimed at tying financial incentives to a variety of accountability indicators including student achievement ("What works," 1997; Bowers, 1989). According to a 1997 report by the Council of Chief State School Officers, 46 states have subsequently established some form of educational accountability system via statewide testing ("Statewide Assessments Nearly Universal," 1997). These initiatives have generated several side effects, both positive and negative: one, there has been a renewed interest in research on factors that influence student achievement; two, the general public, ever-wary of tax increases, has been given a concrete measurement (however controversial) by which to gauge student success; and three, teachers, administrators, and other professional educators have become increasingly aware of the public policy implications that quantitative data can have on schools, personnel, and school programs (Young, 1996).

As the result of a school funding equity lawsuit in the state of Tennessee, new legislation mandating revised school funding formulae and accountability

procedures was implemented in 1991 for all Tennessee K-12 public schools. Included as part of the accountability procedure is mandated annual testing of all students in grades 2 through 8 in the areas of science, math, language arts, reading and social studies. Enacted in 1992, the Tennessee Education Improvement Act (EIA) of 1991 (passed as a result of the "small school systems" lawsuit) reads in part as follows:

> If school districts do not have mean rates of gain equal to or greater than the national norms based upon the TCAP tests (or tests which measure academic performance which are deemed appropriate), each school district is expected to make statistically significant progress toward that goal...Schools or school districts which do not achieve the required rate of progress may be placed on probation as provided in section 49-1-602 of the Tennessee Code Annotated. If national norms are not available, then the levels of expected gain will be set upon the recommendation of the commissioner with the approval of the state board. (Tennessee Code Annotated § 49-1-601c).

The goal of the program is to reduce variability among scores in school systems across the state regardless of SES (equity) and to ensure that all students are progressing (or "adding value") from one year to the next in each of the key subject areas. These goals are reflective of the national trend toward increased accountability in education. An overview of the variety of approaches utilized in identifying effective schools can be found in Darling-Hammond et al. (1991), Westbrook (1987), Hawley et al. (1984), Mace Matluck (1982), Becker (1992), Bullard et al. (1993) and Lezotte (1989, 1993).

Research Questions

Given the background of Tennessee's accountability system and testing procedure, a particularly meaningful area of inquiry was deemed to be a study of test score data to determine if there was, indeed, evidence of more "equity" and

"value-added" in student scores. Was variability in scores decreasing? How do scores compare across years and grade levels? What are the implications of these findings for curriculum and assessment reforms, particularly in the areas of science and math? Science scale score system-level TCAP data for the years for the years 1990-1994 formed the basis for the analysis.

The null hypothesis of the investigation was:

There is no difference in science scale scores across years or grade levels.

Instrumentation

The instrument used as a test of student knowledge for the years 1990-1994 in Tennessee was the Comprehensive Test of Basic Skills, Fourth Edition (CTBS/4) McGraw-Hill test for grades two through eight. The various portions of the Tennessee achievement tests are referred to as TCAP (Tennessee Comprehensive Assessment Program) tests. The science subsection of the CTBS/4 contains 20 items. There is a wealth of research which has been conducted using the various subtests of the CTBS/4 test battery and, in addition to the CTB technical manual, there have been several reviews of the CTBS/4 published (Hopkins, 1992; Miller, 1992). According to Miller the fourth edition reports estimates of internal consistency and has offered a shift in emphasis to more complex objectives. Hopkins, however, is uncertain whether the IRT model resulted in the elimination of test items because of "lack of model fit" that "assess relevant, but idiosyncratic content/skills" (p. 217.). Both reviewers agree that weaknesses in the battery were not restricted to the CTBS/4 but were common to the majority of standardized tests.

Data Analysis and Results

The original data set consisted of scale science scores for the 138 Tennessee public school systems, grades 2-8, for the years 1990 -1994. However, as several systems have only K-6 schools, missing values for those systems reduced the sample size to 133 systems. Initial examination of the data revealed that the greatest range in science scores was in 1990 (min. 624.70 max. 807.30) with a mean of scores in grades 2-8 of 722.22. Science scores in 1994 ranged from a minimum of 625.20 to a maximum of 794.60 with a mean of 725.09 for grades 2-8. (insert Table 1 about here)

Cursory examination did indicate a decrease in dispersion and a slight increase in mean scores across years, except for 1991, the first year under the new accountability plan, which showed a negative change.

Table 2

Mean Science Scale Scores by Year for Grades 2 - 8

1990	1991	1992	1993	1994
722.22	721.42	723.29	723.45	725.09

Table 3

Mean Science Scale Scores Across Years 1990 - 1994

Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
667.51	690.96	713.44	728.49	739.64	754.89	766.82

Using the SPSS for Windows 7 statistical software program, a within-subjects MANOVA was conducted with five levels for year and seven levels for grade, thus creating 35 dependent variables. The within-subjects tests for effect

of YEAR and GRADE led to rejection of the hypothesis of covariance (Mauchly = .0000, signif = .) thus calling into question the robustness of a univariate or mixed model approach without adjustment of numerator and nominator degrees of freedom. However, the Mauchly test is known to be significant for large sample sizes even when the impact of the violation of this assumption is small (Norusis, 1990). Given the large sample size, the determination was made that the violation of this assumption would have little impact on the analysis. The hypothesis that the year, the grade level, and the interaction between year and grade level do not affect science scale scores was rejected. The univariate test for YEAR resulted in an effect size of .216, with the greatest effect for YEAR occurring in 1991. (Insert Table 4 about here).

An examination of the GRADE effect showed a Pillais of .993 and Eta squared for grade 3 of .918, thus after controlling for year, third grade accounted for almost 92% of deviation from constant in the science scores. The grade level effect accounted for over twice the deviation of the year effect, with grades 4, 5, and 7 respectively accounting for most deviation after grade 3 (Eta-squared gr3 = .985, gr4 = .72, gr5 = .54, gr7 = .29). (Insert Table 5 about here).

Univariate tests of the year by grade effect resulted in a partial Eta-squared of .294 resulting in the conclusion that the interaction of YEAR by GRADE is much less powerful an effect than that of grade level. (Insert Table 6 about here).

However, an examination of the scale score means across grade levels suggests that the mean scores for science are somewhat higher for each grade level, indicating that as students progress in school they are, on the average, gaining on or exceeding national norms. Also, SDs were generally higher in grades 2 and 3 than in grades 7 and 8 and the SDs show a decrease each year

as grade level increases, suggesting that the longer students remain in school, the less variation there is in their science scores.

Conclusions

Several conclusions and implications for educational policymakers are clear upon examination of these data. First, it is evident that variability in science achievement decreases as students progress in school. More in-depth study by individual school systems should be undertaken to determine if these results are indicative of a regression to the mean or "floor and ceiling" effect or whether they are indicative of the effects of educational programs in place at school and system level. Several ongoing studies across the state of Tennessee are currently examining the effects of "building change" on student achievement (Sanders et al., 1994; Bobbett et al., 1991). Second, the highest maximum science scale score of the school systems across Tennessee occurred in 1990 in grade 8 (807). In 1991 the maximum science scale score was 788.9, in 1992 the highest maximum science scale score was 795.3, in 1993 the highest maximum science scale score was 794.6, and in 1994 the maximum science scale score was 787.4. While these data possibly bear out the leveling effects of the changes in school funding, school level and system level administration should certainly wish to examine their individual system data to determine to what extent programs and policies have been affected by funding changes.

List of References

Becker, M. (1992). An effective schools primer. (ERIC Document Reproduction Service No. ED 368 044)

Bobbett, G. C., & French, R. L. (1991). Assessing school climate. (ERIC Document Reproduction Service No. ED 340 738)

Bowers, B. C. (1989). State-enforced accountability of local school districts. (ERIC Document Reproduction Service No. ED 309 556)

Bullard, P., & Taylor, B. O. (1993). Making school reform happen. (ERIC Document Reproduction Service No. ED 356 525)

CTB. (Undated). Comprehensive Test of Basic Skills Spring Norms Book: March through June (4th ed.). Monterey, CA: Author.

Darling-Hammond, L., & Ascher, C. (1991). Accountability mechanisms in big city school systems. New York: ERIC Clearinghouse on Urban Education. (ERIC Document Reproduction Service No. ED 334 339)

Gardner, D. P., et al. (1983). A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the Secretary of Education. (ERIC Document Reproduction Service No. ED 226 006)

Hawley, W., Goodstein, H., Rosenholtz, S., & Hasselbring, T. (1984). Good schools: What research says about improving student achievement. Peabody Journal of Education, 61(4), 1-135.

Hopkins, K. D. (1992). Review of the Comprehensive Tests of Basic Skills, Fourth Edition. In Kramer, J. J. & Conoley, J. C. (Eds.) (1992). The eleventh mental measurements yearbook. Lincoln: University of Nebraska, Buros Institute of Mental Measurement.

Lezotte, L. (1989). Base school improvement on what we know about effective schools. American School Board Journal, 176(8), 18-20.

Lezotte, L. W. (1993). Creating effective schools today and tomorrow. Journal for Quality and Participation, 16(1), 22 - 30.

Mace-Matluck, B. (1982). The effective schools movement: Its history and context. An SEDL monograph. (ERIC Document Reproduction Service No. ED 304 781)

Miller, M. D. (1992). Review of the Comprehensive Tests of Basic Skills, Fourth Edition. In Kramer, J. J. & Conoley, J. C. (Eds.) (1992). The eleventh mental measurements yearbook. Lincoln: University of Nebraska, Buros Institute of Mental Measurement.

Norusis, M. J. (1993). SPSS for Windows Base System User's Guide Release 6.0. Chicago: SPSS, Inc.

Sanders, W. L., Saxton, A. M., Schneider, J. F., Dearden, B. L., Wright, S. P., & Horn, S. P. (1994). Effects of building change on indicators of student academic growth. Evaluation Perspectives, 4(1), 3.

Statewide assessments nearly universal, chiefs say. (1997) What Works in Teaching and Learning, 29(10), 1 - 2.

Westbrook, J. D. (1982). Considering the research. What makes an effective school? Southwest Educational Development Lab, Austin, Tex.; National Inst of Education (ED), Washington, DC.

Young, T. A. (1996). A study of the perceptions of Tennessee teachers, principals, superintendents, legislators, and Department of Education representatives regarding the impact of the Tennessee value-added assessment system. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.

Table 1 (n = 133)

Variable	$\bar{x}$	min	max	variance	SD
SS90.2	665.86	624.70	698.70	142.07	11.92
SS90.3	689.95	647.90	716.80	109.29	10.45
SS90.4	706.92	677.70	733.60	104.85	10.24
SS90.5	727.62	702.70	765.90	90.82	9.53
SS90.6	741.98	709.60	774.70	115.11	10.73
SS90.7	756.30	725.90	796.60	113.91	10.67
SS90.8	766.88	737.30	807.30	84.64	9.20
SS91.2	667.56	632.90	728.60	213.08	14.60
SS91.3	689.17	657.50	714.50	108.14	10.40
SS91.4	709.41	663.40	730.90	98.77	9.94
SS91.5	727.18	691.00	748.10	97.49	9.87
SS91.6	740.84	709.90	778.60	89.08	9.44
SS91.7	752.92	719.50	775.30	109.61	10.47
SS91.8	763.42	730.90	788.90	101.44	10.07
SS92.2	667.01	630.70	697.50	155.65	12.48
SS92.3	690.57	662.90	720.30	118.94	10.91
SS92.4	718.57	695.90	739.50	60.55	7.78
SS92.5	727.22	690.90	774.20	81.25	9.01
SS92.6	734.00	699.00	763.90	106.63	10.33
SS92.7	757.62	730.10	781.30	66.33	8.14
SS92.8	768.07	740.80	795.30	90.01	9.49
SS93.2	662.57	627.90	692.90	157.98	12.57
SS93.3	686.48	653.50	717.40	119.89	10.95
SS93.4	716.46	681.60	741.40	119.28	10.92
SS93.5	726.97	699.60	751.20	72.34	8.51
SS93.6	746.42	705.70	775.60	106.53	10.32
SS93.7	754.55	729.70	779.00	61.99	7.87
SS93.8	770.67	747.80	794.60	54.75	7.40
SS94.2	674.56	625.20	714.60	166.74	12.91
SS94.3	698.61	650.10	732.50	162.42	12.75
SS94.4	715.85	682.30	743.60	95.15	9.76
SS94.5	733.48	698.90	754.30	87.22	9.34
SS94.6	734.98	698.20	756.50	79.74	8.93
SS94.7	753.05	720.80	784.30	73.07	8.55
SS94.8	765.08	745.60	787.40	60.60	7.79

Table 4

Effect YEAR

Univariate F-tests with (1,132) D. F.

Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	Sig of F
YR91	5726.305	12282.853	5726.305	93.052	61.539	0.000
YR92	451.511	4771.883	451.511	36.151	12.490	0.001
YR93	129.378	4615.824	129.378	34.968	3.700	0.057
YR94	903.382	4560.191	903.382	34.547	26.150	0.000

Tests involving 'YEAR" Within-Subject Effect

Source	SS	DF	MS	F	Sig of F	Partial ETA Sqd
Within	26230.75	528	49.68			
YEAR	7210.58	4	1802.64	36.29	0.00	0.216

Table 5

Effect GRADE

Univariate F-test with (1,132) D. F.

Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	Sig of F
GRD3	4876066.400	72482.689	4876066.400	549.111	8879.924	0.000
GRD4	84298.429	32228.738	84298.430	244.157	345.263	0.000
GRD5	9901.317	8356.244	9901.317	63.305	156.407	0.000
GRD6	151.906	14257.135	151.906	108.009	1.406	0.238
GRD7	5159.598	12360.349	5159.598	93.639	55.101	0.000

Tests Involving 'GRADE' Within-subject Effect

Source	SS	DF	MS	F	Sig of F	Partial ETA Sqd
Within	145817.56	792	184.11			
GRADE	4975745.57	6	829290.93	4504.25	0	0.972

Table 6

Effect YEAR by GRADE

Univariate F-tests with (1, 132) D. F.

Variable	Hypoth SS	Error SS	Hypoth MS	Error MS	F	Sig of F
Y91G3	4521.112	11371.945	4521.112	86.151	52.479	0.000
Y91G4	62.820	7401.264	62.820	56.070	1.120	0.292
Y91G5	4150.812	7046.591	4150.812	53.383	77.755	0.000
Y91G6	439.102	4660.127	439.102	35.304	12.438	0.001
Y91G7	1457.738	6050.607	1457.738	45.838	31.802	0.000
Y91G8	2.188	4004.411	2.188	30.336	0.072	0.789
Y92G3	3519.216	6509.796	3519.216	49.317	71.360	0.000
Y92G4	411.994	6368.512	411.994	48.246	8.539	0.004
Y92G5	252.361	4613.123	252.361	34.948	7.221	0.008
Y92G6	37.356	3243.232	37.356	24.570	1.520	0.220
Y92G7	2396.705	5344.738	2396.705	40.490	59.192	0.000
Y92G8	4513.599	4336.203	4513.599	32.850	137.400	0.000
Y93G3	11447.664	5048.860	11447.664	38.249	299.294	0.000
Y93G4	433.710	5000.473	433.710	37.882	11.449	0.001
Y93G5	2.548	4353.529	2.548	32.981	0.077	0.781
Y93G6	119.436	4225.042	119.436	32.008	3.731	0.056
Y93G7	21.319	4933.824	21.319	37.377	0.570	0.451
Y93G8	3768.776	4003.734	3768.776	30.331	124.254	0.000
Y94G3	1744.383	3843.347	1744.383	29.116	59.911	0.000
Y94G4	886.185	3644.435	886.185	27.609	32.097	0.000
Y94G5	2077.419	3825.429	2077.419	28.981	71.683	0.000
Y94G6	1032.643	2985.986	1032.643	22.621	45.650	0.000

Tests involving YEAR BY GRADE Within-subjects effects

Source	SS	DF	MS	F	Sig of F	Partial ETA Sqd
Within	120649.26	3168	38.08			
YEAR by GRADE	50211.76	24	2092.16	54.94	0.000	0.294